МОДЕЛЬ НАРУШИТЕЛЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ИНТЕЛЛЕКТУАЛИЗИРОВАННЫХ КИБЕРФИЗИЧЕСКИХ СИСТЕМ КРИТИЧЕСКОЙ ИНФОРМАЦИОННОЙ ИНФРАСТРУКТУРЫ

Максимова Е.А.[1]

Ключевые слова: *искусственный интеллект, деструктивные воздействия, функциональный агент, инфраструктурный деструктивизм, каскадные последствия, автономность, инфраструктурный риск, регулятивный подход, межобъектные зависимости, источник воздействия, деструктивный эффект.*

Аннотация

Цель работы: *развитие модели нарушителя информационной безопасности применительно к интеллектуализированным киберфизическим системам критической информационной инфраструктуры с учетом автономных программно-алгоритмических компонентов, межобъектных зависимостей и каскадного распространения деструктивных эффектов.*

Методы исследования: *системный анализ, структурно-функциональное и теоретико-множественное моделирование, классификация источников и сценариев деструктивных воздействий, элементы риск-ориентированного подхода.*

Результаты исследования: *предложена расширенная модель нарушителя информационной безопасности интеллектуализированных киберфизических систем критической информационной инфраструктуры, учитывающая автономность и адаптивность интеллектуальных компонентов, динамику межсистемных связей и возможность каскадного распространения нарушений. Определены условия перехода системы искусственного интеллекта в функционально-субъектное состояние; выделены антропогенные, программно-алгоритмические, инфраструктурно-технологические и гибридные категории источников и агентов воздействия. Сформулированы признаки инфраструктурно-динамического характера угроз, при которых деструктивный эффект возникает в результате взаимодействия человека, алгоритмов, данных, технических компонентов и текущего состояния инфраструктуры. Формализованы сценарии формирования деструктивного эффекта. Предложена структура показателей инфраструктурного риска, учитывающая автономность, адаптивность и каскадное распространение последствий. Обоснована возможность применения разработанной модели для раннего обнаружения и прогнозирования угроз, локализации нарушений, предупреждения каскадной деградации и выбора механизмов реконфигурации и самовосстановления интеллектуализированных киберфизических систем. Полученные результаты могут применяться при моделировании угроз и обосновании мер защиты интеллектуализированных объектов КИИ.*

DOI: 10.24412/1994-1404-2026-3-7-16

Введение

В сфере информационной безопасности Российской Федерации (ИБ) за последние десятилетия сформировалась устойчивая концептуальная и методологическая база, определяющая объект защиты, источники угроз, модель нарушителя ИБ (далее — Модель), состав мер защиты и порядок оценки рисков. Данный подход закреплён в документах и регламентах в области защиты информации, в том числе рассматривающих вопросы обеспечения безопасности критической информационной инфраструктуры Российской Федерации[2] (КИИ) [1].

[2] Федеральный закон от 26.07.2017 № 187-ФЗ «О безопасности критической информационной инфраструктуры Российской Федерации» // Собрание законодательства Российской Федерации. 2017. № 31. Ст. 4736.

[1] ***Максимова Елена Александровна****, доктор технических наук, доцент, главный инженер-исследователь, Научный центр информационных технологий и искусственного интеллекта, Научно-технологический университет «Сириус», пгт Сириус, Российская Федерация. ORCID 0000-0001-8788-4256.*

E-mail: maksimova.ea@talantiuspeh.ru

В контексте регулятивного (классического) подхода существующие Модели преимущественно описывают характеристики физического лица или группы лиц[3] [2, 3]. При этом автономные программно-алгоритмические компоненты, межобъектные зависимости и способность инфраструктуры к каскадному усилению последствий либо рассматриваются как внешние условия реализации угроз, либо не включаются в структуру Модели. Это затрудняет описание сценариев, в которых источник воздействия, непосредственный агент и механизм распространения деструктивного эффекта контекстно различимы.

Вместе с тем применение классической теории ИБ к КИИ не всегда является достаточным. В КИИ это связано с ее специфическими свойствами как сложной системы [4, 5]. Кроме того, в качестве объектов КИИ сегодня активно рассматриваются киберфизические системы (КФС) [6—11], в интеллектуализированном состоянии способные выступать в роли как объекта защиты, так и субъекта нападения (агента воздействия).

Под интеллектуализированной киберфизической системой (ИКФС) в настоящем исследовании понимается совокупность взаимосвязанных информационных, вычислительных, коммуникационных и физических компонентов, в которой функции анализа состояния, прогнозирования, принятия решений или управления частично либо полностью реализуются с применением методов искусственного интеллекта. В составе ИКФС система искусственного интеллекта (СИИ) может выступать программно-алгоритмическим компонентом, реализующим функции анализа, прогнозирования, принятия решений или управления [12].

В данном исследовании ИКФС будем рассматривать как целостный объект или совокупность взаимосвязанных объектов КИИ, а СИИ — как её интеллектуальный компонент, потенциально способный приобретать свойства функционального агента воздействия (далее — ФАВ).

Инфраструктурный деструктивизм (далее — ИнД) будем рассматривать как способность сложной информационной инфраструктуры порождать, распространять и усиливать деградационные состояния за счёт собственных связей, архитектуры, режимов взаимодействия и автономных компонентов. В отличие от классической модели «нарушитель — уязвимость — атака — ущерб» ИнД позволяет выполнить переход к модели, в которой учитываются не только субъектные действия, но и внутренние свойства инфраструктуры как источника деструктивной динамики[4] [13, 14].

Классическое понимание нарушителя ИБ основывается на представлении о физическом лице или логическом объекте, случайно или преднамеренно совершившем действие, следствием которого является нарушение ИБ организации[2]. В методических документах ФСТЭК России при оценке угроз ИБ акцент сделан на антропогенных источниках угроз. Данный подход является нормативно значимым и практически необходимым, однако применительно к ИКФС, он нуждается в развитии. Это обоснованно, так как он не в полной мере учитывает случаи, когда непосредственный агент воздействия имеет программно-алгоритмическую природу, а деструктивный эффект — инфраструктурный, распределённый или каскадный генезис. В рамках настоящего исследования расширяется не нормативное понятие актуального нарушителя, а состав элементов, учитываемых при построении его модели. Наряду с антропогенным источником угрозы рассматриваются ФАВ, механизм реализации воздействия и инфраструктурное распространение последствий. Такой подход не предполагает признания СИИ самостоятельным субъектом юридической ответственности.

Классификация деструктивных воздействий на элементы КИИ

Для уточнения Модели классифицируем деструктивные воздействия в зависимости от их генезиса и влияния на межобъектные связи по следующим сценариям.

Сценарий 1 соответствует традиционной Модели и связан с проявлением ИнД вследствие целенаправленной деятельности субъекта, которому присваивается роль злоумышленника прямого воздействия[6—8].

Сценарий 2 также частично учитывается действующими методическими подходами и связан с проявлением ИнД вследствие ошибочных действий субъекта, рассматриваемого в качестве нарушителя непреднамеренного воздействия [2, 5, 6, 7].

Сценарий 3 связан с проявлением ИнД вследствие межобъектного взаимодействия. Реализуемые в ходе данного сценария процессы соответствуют представлениям о каскадных отказах, нормальных авариях, функциональном резонансе и взаимозависимости объектов критических инфраструктур.

Сценарий 4 возможен в условиях КИИ, где используемые в КФС СИИ выполняют функции анализа, прогнозирования, принятия решений и формирования управляющих воздействий. При достижении определённого уровня автономности и доступа к критически

---

[3] ФСТЭК России. Методика оценки угроз безопасности информации. Утв. ФСТЭК России 05.02.2021.

[4] Максимова Е.А. Модели и методы оценки информационной безопасности субъекта критической информационной инфраструктуры при деструктивных воздействиях инфраструктурного генеза : дис. … докт. тех. наук: 2.3.6 / Максимова Елена Александровна, 2022. 448 с. EDN: OHDNPO.

[5] ГОСТ Р 53114-2008. Защита информации. Обеспечение информационной безопасности в организации. Основные термины и определения. М. : Стандартинформ, 2009.

[6] ГОСТ Р ИСО/МЭК 27002-2021. Информационная технология. Методы и средства обеспечения безопасности. Свод норм и правил по применению мер обеспечения информационной безопасности. М. : Стандартинформ, 2021.

[7] ГОСТ Р ИСО/МЭК 27005-2010. Информационная технология. Методы и средства обеспечения безопасности. Менеджмент риска информационной безопасности. М. : Стандартинформ, 2011.

значимым функциям такие системы приобретают свойства ФАВ на КИИ. В зависимости от происхождения целевой функции, режима эксплуатации и характера последствий СИИ может рассматриваться как инструмент антропогенного нарушителя, скомпрометированный логический компонент, источник непреднамеренного алгоритмического воздействия либо элемент гибридного распределённого сценария нарушения ИБ. Таким образом данный сценарий связан с автономным алгоритмическим воздействием, включая действия СИИ. То есть СИИ может рассматриваться как автономный агент управления или как катализатор межобъектного деструктивизма[8, 9] [13].

Предложенная система сценариев обосновывает выход за пределы классической теории ИБ и описывает внутреннюю способность сложной системы порождать условия саморазрушения через собственные связи, алгоритмы, архитектуру и режимы взаимодействия.

С учетом предложенных сценариев, деструктивные воздействия на элементы КИИ можно разделить на три класса.

Класс 1: прямые деструктивные воздействия на объекты КИИ и/или на межобъектные связи. Для данного класса применимы действующие процедуры моделирования угроз, оценки возможностей нарушителя, категорирования объектов и выбора мер защиты[3, 10].

Класс 2: косвенные антропогенно обусловленные, как преднамеренные, так и непреднамеренные воздействия на объекты КИИ[12] [15]. Нарушитель выступает в том числе и как источник условий, порождающих уязвимости, ошибки проектирования, дефекты программного кода, конфигурационные нарушения или эксплуатационные несогласованности[5, 6, 11, 12].

В данных ситуациях ИКФС может рассматриваться как объект КИИ, то есть как объект защиты.

Класс 3: косвенные воздействия нарушителя на объекты КИИ инфраструктурного генезиса технологического субъектного типа. Нарушитель является создателем технологического субъекта, способного самостоятельно реализовать деструктивные воздействия инфраструктурного генеза.

По сути дела, на уровне деструктивных воздействий класса 3 возникает много вопросов, связанных с деятельностью ИКФС. Именно третий класс представляет наибольший научный интерес, поскольку он выходит за пределы классической Модели. Если интеллектуальному компоненту ИКФС переданы функции лица, принимающего решение, то фактически она приобретает свойства функционального субъекта воздействия. При этом, оставаясь программным продуктом, СИИ в данном случае может содержать ошибки, алгоритмические погрешности и иные дефекты, способные привести к деструктивному результату[3, 13, 14].

Переход программной системы из объектного в функционально-агентное состояние

Включение СИИ в контуры управления КИИ требует рассмотрения проблемы перехода программной системы в функционально-агентное состояние. В классическом понимании программная система является объектом защиты, инструментом обработки данных или средством автоматизации. Однако при передаче ей функций принятия решений она приобретает признаки функциональной агентности[12, 13, 15, 16, 17—20].

Рассмотрение этого перехода возможно на философском, программно-аппаратном, техническом, формально-математическом, этическом и нормативно-правовом уровнях. Философский уровень позволяет осмыслить природу субъектности, агентности и автономности программных систем. Технический — описывает механизмы реализации управляющих функций. Формально-математический — необходим для моделирования состояний, переходов и последствий решений. Этический и нормативно-правовой — связаны с вопросами ответственности, контроля и допустимых границ автономности[12, 13]. Системообразующим в данном случае будет уровень философского познания, через который возможно принятие сущностей процессов перехода программной системы из объектного в субъектное состояние.

Согласно ГОСТ[4], нарушитель ИБ организации — это физическое лицо или логический объект, случайно или преднамеренно совершивший действия, следствием которых является нарушение ИБ организации. В данном определении уже содержится возможность его рассмотрения как логического объекта. В это же время в Методике нарушители ИБ рассматриваются как антропогенные источники угроз, то есть лица или группы лиц[6].

Таким образом, с одной стороны, ГОСТ[4] допускает рассмотрение логического объекта в качестве нарушителя. С другой стороны, Методика[2] конкретизирует актуального нарушителя преимущественно через

---

[8] ISO/IEC 23894:2023. Information technology — Artificial intelligence — Guidance on risk management. Geneva: ISO, 2023.

[9] ISO/IEC 42001:2023. Information technology — Artificial intelligence — Management system. Geneva: ISO, 2023.

[10] Приказ ФСТЭК России от 21.12.2017 № 235 «Об утверждении Требований к созданию систем безопасности значимых объектов критической информационной инфраструктуры Российской Федерации и обеспечению их функционирования».

[11] ГОСТ Р 56939-2016. Защита информации. Разработка безопасного программного обеспечения. Общие требования. М. : Стандартинформ, 2016.

[12] ГОСТ Р 56545-2015. Защита информации. Уязвимости информационных систем. Правила описания уязвимостей. М. : Стандартинформ, 2016.

[13] ISO/IEC 23894:2023. Information technology — Artificial intelligence — Guidance on risk management. Geneva: ISO, 2023.

[14] ISO/IEC 42001:2023. Information technology — Artificial intelligence — Management system. Geneva: ISO, 2023.

[15] ПНСТ 1046-2026. Искусственный интеллект в критической информационной инфраструктуре. Общие положения.

[16] ГОСТ Р 59276-2020. Системы искусственного интеллекта. Способы обеспечения доверия. Общие положения. М. : Стандартинформ, 2020.

[17] ГОСТ Р 59921.0-2022. Системы искусственного интеллекта. Часть 0. Основные положения. М. : Стандартинформ, 2022.

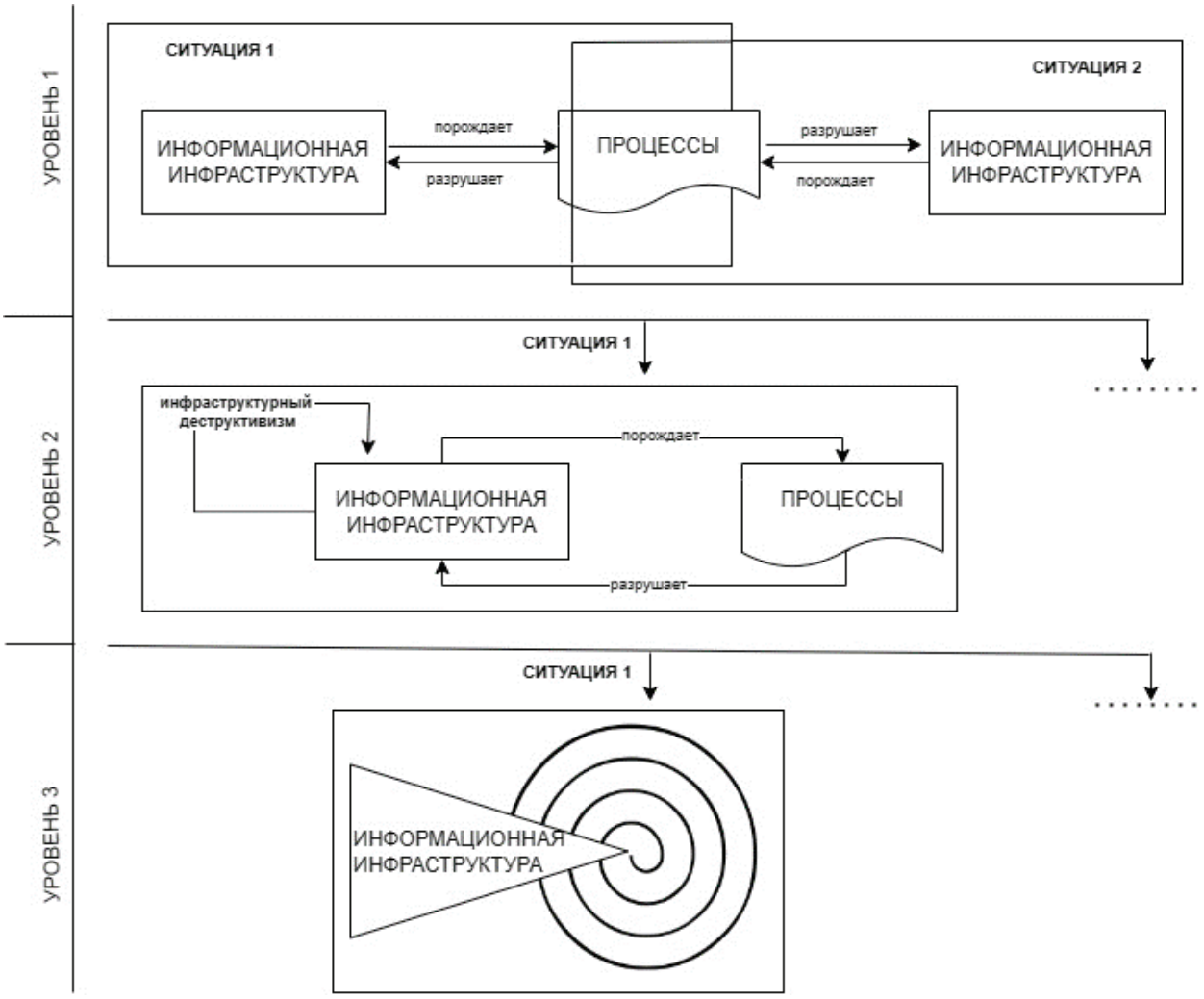


Рис. 1. Трехуровневая модель детализации развития инфраструктурного деструктивизма

физическое лицо или группу лиц. В контексте ИнД это значимо, поскольку ФАВ может быть логическим объектом, функционально реализующим субъектное воздействие на инфраструктуру.

В классической теории нарушитель и объект информатизации являются различными сущностями. В теории ИнД нарушитель и объект воздействия могут частично совпадать, а деструктивный эффект является результатом не внешней атаки, а следствием внутренних связей, алгоритмов и режимов функционирования. Данное явление представимо в виде трехуровневой модели детализации процесса развития ИнД в информационной инфраструктуре (рис. 1).

На первом уровне (рис. 1) отражается взаимное влияние информационной инфраструктуры и порождаемых ею процессов. На втором уровне показано формирование контура положительной обратной связи, в котором процессы, сформированные инфраструктурой, изменяют состояние её компонентов и связей. На третьем уровне представлено накопление и усиление деструктивных изменений, приводящее к переходу инфраструктуры в область преддеструктивных или деструктивных состояний.

Ограничения субъектной парадигмы в контексте инфраструктурного деструктивизма

Описанную ситуацию можно обозначить как кризис субъектной парадигмы. С философской точки зрения субъектность традиционно связывается с активностью, намерением, рефлексией и способностью быть источником действия. Однако в условиях распределённых, автоматизированных и интеллектуализированных инфраструктур деструктивный эффект может возникать без явно выраженного субъекта-инициатора. В этом случае субъект перестаёт быть единственным и обязательным центральным элементом объяснения нарушения.

В традиционной Модели ключевым является вопрос: кто совершил деструктивное воздействие? В инфраструктурной Модели центр внимания смещается от субъекта к структуре, от намерения к зависимости, от события к процессу, от единичного воздействия к динамике распространения последствий.

В КИИ КФС представляют собой совокупность взаимозависимых компонентов. Поэтому нарушение в одном объекте может повлиять на другие объекты и на инфраструктуру в целом. ИнД фиксирует эту

особенность и описывает деструктивность не только как результат субъектного воздействия, но и как системное свойство инфраструктуры, способной порождать, распространять и усиливать деградационные состояния. Таким образом, ИнД можно рассматривать как методологический ответ на кризис субъектной парадигмы в сфере обеспечения безопасности КИИ.

Важно также отметить, что использование в КИИ СИИ дополнительно размывает и усложняет классическое понимание субъекта. СИИ может не обладать субъектностью в человеческом смысле, но при этом она способна принимать решения, инициировать действия и др. В результате возникает ситуация, при которой ФАВ является одновременно объектом защиты, инструментом управления и активным агентом воздействия на инфраструктуру.

Кроме того, если интеллектуализированный компонент КФС, действуя в соответствии с заданным алгоритмом и текущими параметрами среды, перераспределяет нагрузку таким образом, что это приводит к каскадной деградации, деструктивный эффект не может быть полностью объяснён намерением разработчика, ошибкой оператора или отказом отдельного объекта. В этом случае Модель должна учитывать также способность системы к автономному изменению состояния. Деструктивный эффект при этом будет возникать на пересечении данных, алгоритма, архитектуры, текущего состояния инфраструктуры и межобъектных зависимостей[12].

Таким образом, из-за ограниченности традиционной субъектной парадигмы, предлагаемый инфраструктурно-динамический подход рассматривает нарушителя как элемент распределённого процесса воздействия. В рамках данного подхода деструктивность определяется не только намерением конкретного субъекта, но и результатом межфакторного взаимодействия внутри ИКФС.

Интеллектуализированная киберфизическая система как функциональный агент воздействия на объекты КИИ

Теория ИнД предполагает возможность существования искусственного функционального субъекта, самостоятельно принимающего решения на инфраструктурном уровне. В качестве такого субъекта может выступать СИИ[12-16].

В контексте КИИ возникает вопрос: возможно ли рассматривать ИКФС в качестве объекта КИИ? При ответе на данный вопрос выделим три ситуации.

1) ИКФС идентифицирована как объект КИИ и подлежит защите в соответствии с действующей нормативно-правовой и методической базой[1, 2];

2) ИКФС выступает субъектом деструктивного воздействия на объекты КИИ;

3) ИКФС, первоначально рассматриваемая как объект КИИ, переходит в состояние функционального субъекта воздействия.

В ситуации 1 ИКФС рассматривается как компонент инфраструктуры, к которому применяются регулятивные требования. Ситуации 2 и 3 методологически сложнее, так как ФАВ как субъект деструктивного воздействия, в настоящее время недостаточно формализован. Здесь возникает вопрос: можно ли рассматривать ФАВ как одну из категорий актуального нарушителя?

В соответствии с Методикой[2], ИКФС не может быть напрямую отнесена к актуальным нарушителям, поскольку она не является физическим лицом или группой лиц. Однако если рассматривать ФАВ как антропогенно созданного логического объекта, функционально способного воздействовать на информационные ресурсы и компоненты систем и сетей, то появляется основание для расширения интерпретации Модели. В такой трактовке ФАВ может рассматриваться как функциональный субъект инфраструктурного воздействия.

В расширенной Модели источник деструктивного воздействия не всегда совпадает с его непосредственным функциональным агентом. В инфраструктурных сценариях деструктивный эффект может усиливаться и распространяться межобъектными связями уже после окончания воздействия.

Т.о., в расширенной Модели целесообразно разделять актуального нарушителя в нормативном смысле и ФАВ. Роль актуального нарушителя сохраняется за физическим лицом или группой лиц, выступающих антропогенным источником угрозы. Созданный, настроенный либо скомпрометированный ими логический объект, включая интеллектуальный компонент ИКФС, может рассматриваться как функциональный агент, непосредственно реализующий, адаптирующий или усиливающий воздействие на объекты КИИ. Такое разграничение позволяет включить СИИ в структуру Модели, не наделяя техническую систему правосубъектностью.

Формализация расширенной модели нарушителя информационной безопасности КИИ

Предлагаемая формализация носит концептуальный характер и предназначена для определения состава и взаимосвязи параметров Модели.

В общем виде формализация расширенной Модели в КИИ, с учетом эффектов ИнД, представима в виде:

$$M_н = \langle Src, G, Acc, Res, Act, Aut, Adp, Otar, Ch, Obs, Cas, Rec \rangle, \quad (1)$$

где Src — источник воздействия; G — цели или функциональные приоритеты; Acc — уровень доступа; Res — доступные ресурсы; Act — множество возможных действий; Aut — степень автономности; Adp — способность адаптировать стратегию поведения; Otar — множество целевых объектов; Ch — каналы воздействия; Obs — доступная агенту информация о состоянии инфраструктуры; Cas — потенциал каскадного распространения последствий; Rec — способность к реконфигурации и продолжению воздействия при изменении среды.

Важно отметить, что именно параметры, определяющие способность к адаптивному и автономному поведению, отличают предлагаемую модель от традиционных Моделей.

Модель (1) применяется совместно с моделью КИИ;

$$KII = \langle O, L, S, Act, V, T, R, C, D, E\rangle, \quad (2)$$

где $O = \{o_1, o_2, \ldots, o_n\}$ — множество объектов КИИ; $L = \{l_{ij}\}$ — множество межобъектных связей; $S = \{s_1, s_2, \ldots, s_m\}$ — множество антропогенных субъектов и функциональных агентов воздействия; $Act = \{a_1,a_2,\ldots,a_k\}$ — множество действий и воздействий, способных изменять состояния объектов и связей КИИ; $V = \{v_1, v_2, \ldots, v_p\}$ — множество уязвимостей, дефектов, ошибок конфигурации, архитектурных и алгоритмических несогласованностей; $T = \{t_1, t_2, \ldots, t_r\}$ — множество угроз ИБ; $R = \{r_1, r_2, \ldots, r_q\}$ — множество негативных последствий; $C = \{c_1, c_2, \ldots, c_z\}$ — множество ограничений; $D = \{d_1, d_2, \ldots, d_w\}$ — множество деструктивных состояний инфраструктуры; $E=\{e_1,e_2,\ldots,e_u\}$ — множество условий функционирования инфраструктуры.

С точки зрения регулятивного подхода событие ИБ обычно описывается через воздействие субъекта s_i на объект защиты o_j, через реализуемое им воздействие a_k:

$$s_i \rightarrow a_k \rightarrow o_j \rightarrow r_q, \quad (3)$$

где r_q — негативное последствие нарушения ИБ.

В контексте ИнД схема (3) может быть расширена:

$$\langle s_i, a_k, o_i, l_{ij}, o_j, v_p, c_z\rangle \rightarrow d_w \rightarrow r_q. \quad (4)$$

Здесь удобно рассмотреть показатель деструктивности $I_D(t)$:

$$I_D(t) = F(O(t), L(t), S(t), Act(t), V(t), C(t), E(t)), \quad (5)$$

где E(t) — внешние и внутренние условия функционирования инфраструктуры.

Инфраструктура переходит из нормального состояния в деструктивное, если:

$$I_D(t) > \delta \Rightarrow KII(t) \in \Omega_D, \quad (6)$$

где δ — допустимый порог; Ω_D — область деструктивных состояний КИИ.

В области функционирования инфраструктуры можно выделить области:

1) нормального функционирования — Ω_N,
2) преддеструктивных состояний — Ω_P,
3) деструктивных состояний — Ω_D,
4) восстановления — Ω_R.

Переходы между ними определяют динамику состояний КИИ:

$$\Omega_N \rightarrow \Omega_P \rightarrow \Omega_D \rightarrow \Omega_R \rightarrow \Omega_N.$$

Важно отметить, что в сложноструктурированной инфраструктуре данные переходы не всегда являются линейными.

В контексте ИнД можно ввести показатель R_I — инфраструктурный риск, связанный с участием СИИ в развитии управляющих решений на уровне КИИ:

$$R_I = P(T) \times U \times K_I \times K_c \times K_r, \quad (7)$$

где P(T) — вероятность реализации угрозы; U — величина потенциального ущерба; K_I — коэффициент межобъектной связности; K_c — коэффициент критичности затронутых функций; K_r — коэффициент каскадного распространения последствий.

Данный показатель характеризует инфраструктурную составляющую риска и позволяет учитывать факторы межобъектной связности, критичности функций и каскадного распространения последствий, которые могут не отражаться в локальной объектно-ориентированной оценке риска.

Для случаев, связанных с применением СИИ, модель (7) должна быть дополнена коэффициентом алгоритмической автономности K_ai и коэффициентом доверия к данным K_data:

$$R_AI = P(T_AI) \times U \times K_I \times K_c \times K_r \times K_aut \times K_adp \times K_d,$$

где R_AI — риск, связанный с включенностью ИКФС в управленческие функции на уровне КИИ; P(T_AI) — вероятность реализации угрозы или ошибочного алгоритмического воздействия, связанного с ИКФС; K_aut — коэффициент автономности СИИ; K_adp — коэффициент адаптивности СИИ, характеризующий её способность изменять стратегию поведения в зависимости от данных и состояния инфраструктуры; K_d — коэффициент риска. Увеличение K_d соответствует увеличению риска, связанного со снижением качества данных.

Предложенные показатели имеют концептуальный характер. Значения порогов автономности, адаптивности и доступа, а также коэффициентов связности, критичности и каскадного распространения должны определяться применительно к архитектуре конкретной ИКФС. Практическое использование модели предполагает предварительную нормализацию и калибровку коэффициентов.

Интересна ситуация, в которой СИИ, входящая в состав ИКФС, приобретает свойства ФАВ, т. е. становится потенциальным нарушителем ИБ. В данном случае воздействие ИКФС на КИИ можно представить в виде:

$$Y_{AI}(t) = \Phi_{AI}(X(t), G, H, C_{AI}), \quad (8)$$

где $Y_{AI}(t)$ — управляющее воздействие, сформированное СИИ в момент времени t; X(t) — входные данные; G — целевая функция; H — накопленная история обучения и эксплуатации; C_{AI} — ограничения, правила безопасного функционирования и допустимые режимы принятия решений.

При этом переход ИКФС в функционально-субъектное состояние может быть задан условием:

$$a_{AI} \in S_{FAV}, \text{ если } Aut > \alpha \wedge Acc \geq \beta \wedge Adp \geq \eta \wedge \gamma, \quad (9)$$

где a_{AI} — СИИ, рассматриваемая как потенциальный ФАВ; S_{FAV} — множество ФАВ; α — порог автономности; β — порог доступа к критически значимым функциям; η — порог адаптивности; γ — минимально значимое изменение состояния КИИ, вызванное решением СИИ.

Таким образом, СИИ приобретает свойства ФАВ, если обладает достаточным уровнем автономности и адаптивности, имеет доступ к значимым функциям или объектам КИИ и способна формировать решения, вызывающие существенное изменение состояния инфраструктуры.

Для описания ФАВ, как потенциального нарушителя для объектов КИИ, определим следующие виды сценариев воздействия:

sc_1 — прямое преднамеренное воздействие субъекта;

sc_2 — непреднамеренное ошибочное воздействие субъекта;

sc_3 — межобъектное инфраструктурное воздействие;

sc_4 — автономное или полуавтономное алгоритмическое воздействие ФАВ.

Для каждого сценария можно задать обобщённую функцию деструктивного эффекта:

$$D_sc[i] = F_sc[i](S, O, L, Act, V, C, E), \quad (10)$$

где D_sc[i] — деструктивный эффект в рамках сценария sc[i], $i=\overline{1,4}$; F_sc[i] — функция преобразования исходных условий и воздействий в деструктивный результат. Модели реализации обозначенных видов сценариев представлены в таблице 1.

Таблица 1

Модели реализации сценариев деструктивного воздействия на объекты КИИ

СЦЕНАРИЙ	МОДЕЛЬ	ПОЯСНЕНИЯ
sc_1	$D_sc_1 = F_1(S_h \cup S_g, A, O, V)$	S_h — физическое лицо
sc_2	$D_sc_2 = F_2(S_h \cup S_g, A_err, O, V, C)$	A_err — множество ошибочных действий, включая ошибки эксплуатации, проектирования, настройки, сопровождения или управления. S_g — группа лиц
sc_3	$D_sc_3 = F_3(O, L, V, C, E)$	—
sc_4	$D_sc_4 = F_4(S_I, X, M_{AI}, G, Y, O, L, C)$	S_I — логический объект, выполняющий функции агента воздействия. X — входные данные СИИ. G — целевая функция. Y — сформированное решение или управляющее воздействие.

Таким образом, ИнД может быть формально представлен как динамическое состояние, при котором возможно превышение допустимого порога деструктивности:

$$F(O(t), L(t), S(t), A(t), V(t), C(t), E(t)) > \delta. \quad (11)$$

Данное условие определяет отличие ИнД от классического нарушения ИБ, так как в инфраструктурной модели нарушение может быть результатом сложной динамики инфраструктуры.

Модель является исходной основой для разработки технологии самовосстановления КФС. Она позволяет учитывать факторы инфраструктурной динамики, которые могут не отражаться в объектно-ориентированной оценке рисков ИБ.

Модель актуального нарушителя КИИ с учетом интеллектуализированных киберфизических систем

Для построения Модели в КИИ с учетом ИКФС целесообразно сохранить общую логику Методики[2], но дополнить её параметрами, отражающими алгоритмическую и инфраструктурную природу воздействия. В частности, к исходным данным здесь должны также относиться сведения о функциях ФА, границах её автономности, типах принимаемых решений, источниках данных, архитектуре модели, режимах обучения, процедурах валидации и связях с объектами КИИ[12-16].

Если ФАВ разрабатывается внешним исполнителем, то Модель может быть многокомпонентной. Конкретная структура при этом будет зависеть от сложности инфраструктуры субъекта КИИ, количества задач, решаемых с помощью интеллектуальных систем, уровня автономности ФА, квалификации разработчиков и характера интеграции в контуры управления.

Особое значение имеет вопрос целеполагания. В Методике[2] нарушители признаются актуальными, если возможные цели реализации угроз могут привести к негативным последствиям и соответствующим рискам. В отношении физического лица цель может быть связана с намерением, мотивом или интересом. В отношении ФАВ цель должна пониматься функционально.

В этой связи можно выделить целевые уровни субъектов деструктивного воздействия:

1) субъект-разработчик или оператор ФАВ, выступающий как нарушитель или злоумышленник в классическом смысле.

2) сам ФАВ как функциональный субъект воздействия, выступающий как непреднамеренный нарушитель, как инструмент злоумышленника или как автономный агент, реализующий деструктивный сценарий в пределах заданных алгоритмических правил. Здесь ФАВ может совмещать роли, так как его поведение может быть результатом наложения целевой функции, ошибочных данных, внешних команд и текущего состояния инфраструктуры. Поэтому для ФАВ возможны

состояния: цель непреднамеренного нарушения, цель злоумышленного использования или комбинированная цель, при которой один и тот же алгоритмический процесс приводит к разным типам деструктивных эффектов.

Так как при рассмотрении уровней возможностей ФАВ должны учитываться не только доступ и ресурсы, но и степень автономности, масштаб управляющих полномочий, качество данных, возможность самообучения, уровень интеграции с объектами КИИ, доступ к критическим функциям и наличие ограничителей безопасного поведения, то можно утверждать о необходимости адаптации и расширении уровней возможностей нарушителя, представленных в Методике[2], при рассмотрении ИКФС.

Заключение

В исследовании представлены взаимосвязанные классификации: по природе источника или функционального агента; по характеру воздействия; по сценарию формирования деструктивного эффекта. Представлены классификации источников и агентов, гибридных структур, видов воздействий, сценариев.

В работе предложено расширение Модели применительно к ИКФС КИИ. Показано, что в таких системах исходный источник деструктивного воздействия и функциональный агент могут различаться. Кроме того, обосновано, что результаты воздействий могут формироваться за счёт межобъектных зависимостей, автономных алгоритмических решений и каскадной динамики инфраструктуры.

Представлены отличия между антропогенными, программно-алгоритмическими, инфраструктурно-технологическими и гибридными категориями источников и агентов воздействия. В качестве условий, при которых СИИ, входящая в состав ИКФС, приобретает свойства ФАВ определены достижения порогового уровня автономности и адаптивности, наличие доступа к функциям управления, способность самостоятельно формировать и реализовывать действия, изменяющие состояния объектов КИИ.

Формализованы сценарии формирования деструктивного эффекта. С учетом каскадного усиления последствий, автономности и адаптивности программно-алгоритмических компонентов описана структура оценки инфраструктурного риска.

Полученные в ходе исследования результаты возможно использовать при разработке моделей угроз, уточнении характеристик актуальных нарушителей, выявлении ФАВ и обосновании мер защиты интеллектуализированных объектов КИИ. Развитие предложенной тематики можно направить на апробацию модели на типовых архитектурах киберфизических систем и исследование процессов самовосстановления ИКФС в условиях ИнД.

Исследование выполнено за счет гранта Российского научного фонда № 25-11-20037 «Технология самовосстановления киберфизических систем национальной критической информационной инфраструктуры в условиях беспрецедентного роста угроз безопасности на основе искусственного интеллекта» (руководитель — д.т.н. Петренко С.А.).

Литература

1. Яковлев А.В., Пеливан М.А. Предметно-смысловые ограничения при моделировании компьютерного противоборства // Правовая информатика. 2024. № 3. С. 86—95. DOI: 10.24682/1994-1404-2024-3-86-95. EDN: JYEUMQ.
2. Чечулин А.А. Моделирование нарушителя, инфраструктуры и атак в системах информационной безопасности // Научно-аналитический журнал «Вестник Санкт-Петербургского университета Государственной противопожарной службы МЧС России». 2024. № 2. С. 70—79. DOI: 10.61260/2218-130X-2024-2-70-79. EDN: VHHYHE.
3. Трещев И.А., Монастырная Е.И. Событийная формальная модель поведения злоумышленника // Ученые записки Комсомольского-на-Амуре государственного технического университета. 2024. № 1 (73). С. 42—46. EDN: IZQRGT.
4. Nickolov E. Critical Information Infrastructure Protection Analysis Evaluation and Expectations // Information and Security. An International Journal, vol. 17, 2005, pp. 105—119. URL: https://www.researchgate.net/publication/228343373_Critical_Information_Infrastructure_Protection_Analysis_Evaluation_and_Expectations
5. Sunyaev A. (2024). Critical Information Infrastructures. In: Internet Computing. Springer, Cham. DOI: 10.1007/978-3-031-61014-1_11.
6. Балябин А.А., Петренко С.А. Метод обеспечения киберустойчивости платформ управления киберфизическими объектами на основе кибериммунитета // Защита информации. Инсайд. 2025. № 6 (126). С. 21—29. EDN: ZZTXGC.
7. Соболев Н.В. Систематизация атак с нулевой динамикой для различных типов киберфизических систем // Проблемы информационной безопасности. Компьютерные системы. 2026. № S2 (70). С. 81—95. DOI: 10.66424/2071-8217-2026-S2-8. EDN: KDDCBB.

8. Кравец А.Г., Сальникова Н.А., Долгова Е.П. Анализ методов решения проблем управления киберфизическими системами // Математические методы в технологиях и технике. 2025. № 3. С. 102—106. EDN: VQPCFW.
9. Петренко С.А. Кибериммунология. СПб. : Афина, 2021. 239 с. ISBN 978-5-9909868-7-9. EDN: ARCUWY.
10. Furrer F.J. (2022). Cyber-Physical Systems. In: Safety and Security of Cyber-Physical Systems. Springer Vieweg, Wiesbaden. DOI: 10.1007/978-3-658-37182-1_2.
11. Paul P.K., Chatterjee R., Kayyali M., Das N. (2025). Cyber-Physical Systems: Its Foundation, Emergence, Possible Applications, and Issues with Reference to the Education Sector. In: Choudhury A., Kaushik K., Kumar V., Singh B.K. (eds), Cyber-Physical Systems Security. Studies in Big Data, vol. 154. Springer, Singapore. DOI: 10.1007/978-981-97-5734-3_1.
12. Привалов А.Н., Богомолов А.В., Ларкин Е.В. Моделирование надежности программных компонентов киберфизических систем // Программные продукты и системы. 2025. № 1. С. 47—54. DOI: 10.15827/0236-235X.149.047-054. EDN: KSYSDN.
13. Максимова Е.А. Методы выявления и идентификации источников деструктивных воздействий инфраструктурного генеза // Электронный сетевой политематический журнал «Научные труды КубГТУ». 2022. № 2. С. 86—99. EDN: KTYQJE.
14. Romashkina N.P., Markov A.S., Stefanovich D.V. Information Technologies and International Security. Moscow : IMEMO, 2023. 111 p. URL: https://www.imemo.ru/publications/info/information-technologies-
15. Максимова Е.А., Садовникова Н.П. Межсубъектное взаимодействие как источник деструктивных воздействий на субъекте критической информационной инфраструктуры // Прикаспийский журнал: управление и высокие технологии. 2021. № 2 (54). С. 71—80. DOI: 10.21672/2074-1707.2021.53.1.071-0800. EDN: MJJRWJ.

SECTION:
SYSTEM ANALYSIS, MANAGEMENT, ARTIFICIAL INTELLIGENCE, AND INFORMATION PROTECTION

AN INFORMATION SECURITY ADVERSARY MODEL FOR INTELLIGENT CYBER-PHYSICAL SYSTEMS IN CRITICAL INFORMATION INFRASTRUCTURE

Maksimova E.A.[18]

Keywords: *artificial intelligence, destructive impacts, functional agent, infrastructure destructivism, cascading consequences, autonomy, infrastructure risk, regulatory approach, inter-object dependencies, source of impact, destructive effect.*

Abstract

Purpose of work: *development of a model of an information security adversary in relation to intelligent cyber-physical systems of critical information infrastructure, taking into account autonomous software and algorithmic components, inter-object dependencies, and the cascading propagation of destructive effects.*

Research methods: *system analysis, structural-functional and set-theoretic modeling, classification of sources and scenarios of destructive impacts, elements of a risk-based approach.*

Results of the study: *An extended model of an information security threat actor targeting intelligent cyber-physical systems within critical information infrastructure is proposed; the model accounts for the autonomy and adaptability of intelligent components, the dynamics of inter-system connections, and the potential for the cascading propagation of security breaches. Conditions for an artificial intelligence system to transition into a "functional-agent" state are defined, and categories of impact sources and agents — classified as anthropogenic, software-algorithmic, infrastructural-technological, and hybrid — are identified. Characteristics of infrastructure-dynamic threats are formulated, wherein a destructive effect arises from the interaction between humans, algorithms, data, technical components, and the infrastructure's current state. Scenarios leading to such destructive effects are formalized. A framework of infrastructure risk indicators is proposed, incorporating factors such as autonomy, adaptability, and the cascading propagation of consequences. The applicability of the developed model is substantiated for early threat detection and forecasting, breach containment, the prevention of cascading degradation, and the selection of reconfiguration and self-healing mechanisms for intelligent cyber-physical systems. The results obtained can be utilized in threat modeling and in justifying protection measures for intelligent CII assets.*

[18] ***Elena A. Maksimova****, Dr.Sc. (Technology), Chief Research Engineer, Scientific Center for Information Technology and Artificial Intelligence, Sirius University of Science and Technology, Sirius urban-type settlement, Russian Federation. ORCID: 0000-0001-8788-4256.*
E-mail: maksimova.ea@talantiuspeh.ru

References

1. Iakovlev A.V., Pelivan M.A. Predmetno-smyslovye ogranicheniia pri modelirovanii komp'iuternogo protivoborstva // Pravovaia informatika. 2024. No. 3. S. 86—95. DOI: 10.24682/1994-1404-2024-3-86-95. EDN: JYEUMQ.
2. Chechulin A.A. Modelirovanie narushitelia, infrastruktury i atak v sistemakh informatsionnoi bezopasnosti // Nauchno-analiticheskii zhurnal "Vestnik Sankt-Peterburgskogo universiteta Gosudarstvennoi protivopozharnoi sluzhby MChS Rossii". 2024. No. 2. S. 70—79. DOI: 10.61260/2218-130X-2024-2-70-79. EDN: VHHYHE.
3. Treshchev I.A., Monastyrnaia E.I. Sobytiinaia formal'naia model' povedeniia zloumyshlennika // Uchenye zapiski Komsomol'skogo-na-Amure gosudarstvennogo tekhnicheskogo universiteta. 2024. No. 1 (73). S. 42—46. EDN: IZQRGT.
4. Nickolov E. Critical Information Infrastructure Protection Analysis Evaluation and Expectations // Information and Security. An International Journal, vol. 17, 2005, pp. 105—119. URL: https://www.researchgate.net/publication/228343373_Critical_Information_Infrastructure_Protection_Analysis_Evaluation_and_Expectations
5. Sunyaev A. (2024). Critical Information Infrastructures. In: Internet Computing. Springer, Cham. DOI: 10.1007/978-3-031-61014-1_11.
6. Baliabin A.A., Petrenko S.A. Metod obespecheniia kiberustoichivosti platform upravleniia kiberfizicheskimi ob"ektami na osnove kiberimmuniteta // Zashchita informatsii. Insaid. 2025. No. 6 (126). S. 21—29. EDN: ZZTXGC.
7. Sobolev N.V. Sistematizatsiia atak s nulevoi dinamikoi dlia razlichnykh tipov kiberfizicheskikh sistem // Problemy informatsionnoi bezopasnosti. Komp'iuternye sistemy. 2026. No. S2 (70). S. 81—95. DOI: 10.66424/2071-8217-2026-S2-8. EDN: KDDCBB.
8. Kravets A.G., Sal'nikova N.A., Dolgova E.P. Analiz metodov resheniia problem upravleniia kiberfizicheskimi sistemami // Matematicheskie metody v tekhnologiiakh i tekhnike. 2025. No. 3. S. 102—106. EDN: VQPCFW.
9. Petrenko S.A. Kiberimmunologiia. SPb. : Afina, 2021. 239 s. ISBN 978-5-9909868-7-9. EDN: ARCUWY.
10. Furrer F.J. (2022). Cyber-Physical Systems. In: Safety and Security of Cyber-Physical Systems. Springer Vieweg, Wiesbaden. DOI: 10.1007/978-3-658-37182-1_2.
11. Paul P.K., Chatterjee R., Kayyali M., Das N. (2025). Cyber-Physical Systems: Its Foundation, Emergence, Possible Applications, and Issues with Reference to the Education Sector. In: Choudhury A., Kaushik K., Kumar V., Singh B.K. (eds), Cyber-Physical Systems Security. Studies in Big Data, vol. 154. Springer, Singapore. DOI: 10.1007/978-981-97-5734-3_1.
12. Privalov A.N., Bogomolov A.V., Larkin E.V. Modelirovanie nadezhnosti programmnykh komponentov kiberfizicheskikh sistem // Programmnye produkty i sistemy. 2025. No. 1. S. 47—54. DOI: 10.15827/0236-235X.149.047-054. EDN: KSYSDN.
13. Maksimova E.A. Metody vyiavleniia i identifikatsii istochnikov destruktivnykh vozdeistvii infrastrukturnogo geneza // Elektronnyi setevoi politematicheskii zhurnal "Nauchnye trudy KubGTU". 2022. No. 2. S. 86—99. EDN: KTYQJE.
14. Romashkina N.P., Markov A.S., Stefanovich D.V. Information Technologies and International Security. Moscow : IMEMO, 2023. 111 p. URL: https://www.imemo.ru/publications/info/information-technologies-
15. Maksimova E.A., Sadovnikova N.P. Mezhsub"ektnoe vzaimodeistvie kak istochnik destruktivnykh vozdeistvii na sub"ekte kriticheskoi informatsionnoi infrastruktury // Prikaspiiskii zhurnal: upravlenie i vysokie tekhnologii. 2021. No. 2 (54). S. 71—80. DOI: 10.21672/2074-1707.2021.53.1.071-0800. EDN: MJJRWJ.